Getting things straight with Mahalanobis distance

Vassili Korotkine
Department of Mechanical Engineering, McGill University
817 Sherbrooke Street West, Montreal QC H3A 0C3

August 17, 2023

Contents

1 Mahalanobis distance

This summary document is based on both Bar Shalom's book [1] and the MECH600 Mahalanobis distance slides. An estimator attempts to characterize the distribution on $\mathbf{e}$ as

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}), \tag{1}$$

where $\mathbf{P}$ is the predicted covariance. To assess the estimator consistency, the squared Mahalanobis distance, otherwise known as Normalized Estimate Error Squared (NEES), is defined as

$$d^2 = \mathbf{e}^\mathsf{T}\mathbf{P}^{-1}\mathbf{e}. \tag{2}$$

The square root of the NEES is called Mahalanobis distance, d. For a variable $\mathbf{e}$ that is indeed characterized by covariance $\mathbf{P}$, the Mahalanobis distance will have a defined mean and upper and lower bounds with some confidence bounds. By comparing these to those computed through Monte Carlo simulation, estimator performance is assessed.

1.1 Change of variables

Define the variable $\mathbf{u}$ as

$$\mathbf{u} = \mathbf{L}\mathbf{e}, \tag{3}$$

where $\mathbf{L}$ is the result of a Cholesky factorization, $\mathbf{P} = \mathbf{L}\mathbf{L}^\mathsf{T}$ Then, the NEES may be rewritten as

$$
\begin{aligned}
d^2 &= \mathbf{u}^\mathsf{T}\mathbf{L}^\mathsf{T}\mathbf{P}^{-1}\mathbf{L}\mathbf{u} && (4)\\
&= \mathbf{u}^\mathsf{T}\mathbf{L}^\mathsf{T}\mathbf{L}^{-\mathsf{T}}\mathbf{L}^{-1}\mathbf{L}\mathbf{u} && (5)\\
&= \mathbf{u}^\mathsf{T}\mathbf{u}. && (6)
\end{aligned}
$$

Furthermore, the covariance on $\mathbf{u}$ is identity since

$$
\begin{aligned}
\mathrm{E}[\mathbf{u}\mathbf{u}^\mathsf{T}] &= \mathrm{E}[\mathbf{L}^{-1}\mathbf{e}\mathbf{e}^\mathsf{T}\mathbf{L}^{-\mathsf{T}}] && (7)\\
&= \mathbf{L}^{-1}\mathrm{E}[\mathbf{e}\mathbf{e}^\mathsf{T}]\mathbf{L}^{-\mathsf{T}} && (8)\\
&= \mathbf{L}^{-1}\mathbf{P}\mathbf{L}^{-\mathsf{T}} && (9)\\
&= \mathbf{L}^{-1}\mathbf{L}\mathbf{L}^\mathsf{T}\mathbf{L}^{-\mathsf{T}} && (10)\\
&= \mathbf{1}. && (11)
\end{aligned}
$$

The punchline is that, for an error characterized by $\mathbf{P}$, the NEES is actually equal to

$$
\begin{aligned}
d^2 &= \mathbf{u}^\mathsf{T}\mathbf{u} && (12)\\
&= \sum_{i=1}^{k} u_i^2, \quad u_i \sim \mathcal{N}(0,1). && (13)
\end{aligned}
$$

The sum of k *independent squares of Gaussian variables with unity variance* is the the well-known chi-squared distribution.

1.2 Mean

The mean of the chi-squared distribution with k degrees of freedom can be obtained as

$$
\begin{aligned}
\mathrm{E}[\mathbf{u}^\mathsf{T}\mathbf{u}] &= \mathrm{E}[\mathrm{tr}(\mathbf{u}^\mathsf{T}\mathbf{u})] && (14)\\
&= \mathrm{E}[\mathrm{tr}(\mathbf{u}\mathbf{u}^\mathsf{T})] && (15)\\
&= \mathrm{tr}(\mathrm{E}[\mathbf{u}\mathbf{u}^\mathsf{T}]) && (16)\\
&= \mathrm{tr}(\mathbf{1}_{k\times k}) && (17)\\
&= k. && (18)
\end{aligned}
$$

where the trace properties were used

$$
\begin{aligned}
\mathrm{tr}(\mathbf{A}\mathbf{B}) &= \mathrm{tr}(\mathbf{B}\mathbf{A}), && (19)\\
\mathrm{E}[\mathrm{tr}(\mathbf{A})] &= \mathrm{tr}(\mathrm{E}[\mathbf{A}]), && (20)
\end{aligned}
$$

where (19) is the cyclic property of the trace and (20) follows from the linearity of the expectation and trace operators.

1.3 Bounds

The cumulative distribution function (CDF) of a probability distribution is given by

$$f_{\text{CDF}}(x) = \int_{-\infty}^{x} p(\tau)\mathrm{d}\tau. \tag{21}$$

Given a value $\bar{x}$, $f_{\text{CDF}}(\bar{x})$ gives the probability that x is less than $\bar{x}$,

$$f_{\text{CDF}}(\bar{x}) = P(x \leq \bar{x}). \tag{22}$$

The CDF may be inverted to give the Probability Point Function (PPF),

$$f_{\text{PPF}}(p) = f_{\text{CDF}}^{-1}(x). \tag{23}$$

Given a probability $\bar{P}$, the PPF gives the value $\bar{x}$ such that x is less than $\bar{x}$ with probability $\bar{P}$,

$$\bar{x} = f_{\text{PPF}}(\bar{P}) \quad \text{s.t. } \bar{P} = P(x \leq \bar{x}). \tag{24}$$

The probability $\bar{P}$ is termed the *confidence threshold*.

The CDF and PPF functions are tabulated for common distributions, including the χ^2 distribution. They may be computed, for instance, with Python's scipy package. Since the Mahalanobis distance for any estimator run should follow the χ^2_k distribution, the upper bound with a given confidence interval may be determined using the CDF and PPF of the χ^2_k distribution.

The lower bound may similarly be determined.

1.4 Single-sided versus double-sided tests

Consider a 95% confidence interval. A one sided probability region is obtained by cutting off a single side of the probability distribution. For the upper bound case this will compute $\bar{x} = f_{\text{PPF}}(0.95)$, which is a $\bar{x}$ such that $x \leq \bar{x}$ with 95% confidence.

The two sided probability region is obtained by cutting off both sides of the probability distribution. Two bounds are computed at values $\bar{x}_{\text{min}} = f_{\text{PPF}}(0.025)$ and $\bar{x}_{\text{max}} = f_{\text{PPF}}(0.975)$ such that $\bar{x}_{\text{min}} \leq x \leq \bar{x}_{\text{max}}$ with 95% confidence. The single and double-sided χ^2_k bounds are illustrated in Figures 1 and 2 respectively.

1.5 Monte Carlo Trials

In a Monte Carlo trials, N independent algorithm runs are conducted, each of which output sequences

$$\{\mathbf{e}_{i,n} \in \mathbb{R}^k, \mathbf{P}_{i,n} \in \mathbb{R}^{k \times k}\}, \tag{25}$$

where $\mathbf{e}_{i,n}$ is the error at the i^{th} timestep and n^{th} Monte Carlo trial, computed from the estimate and ground truth and $\mathbf{P}_{i,n}$ is the covariance output by the estimator. Then the sum of the squared

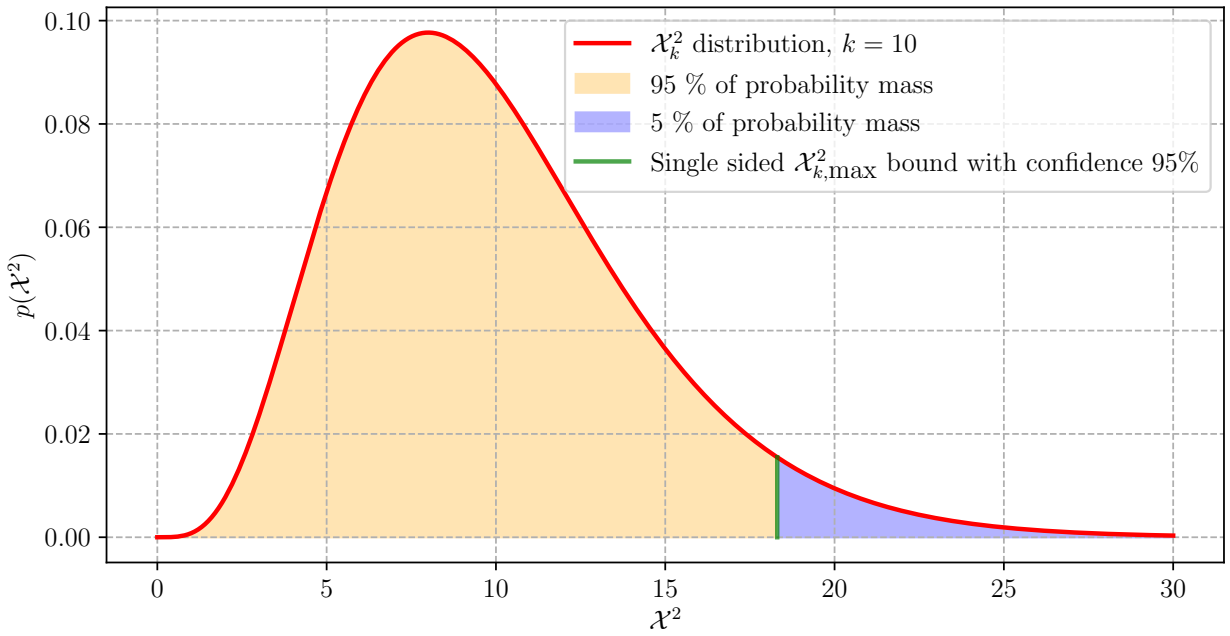

Figure 1: Illustration of single-sided χ^2_k squared test.

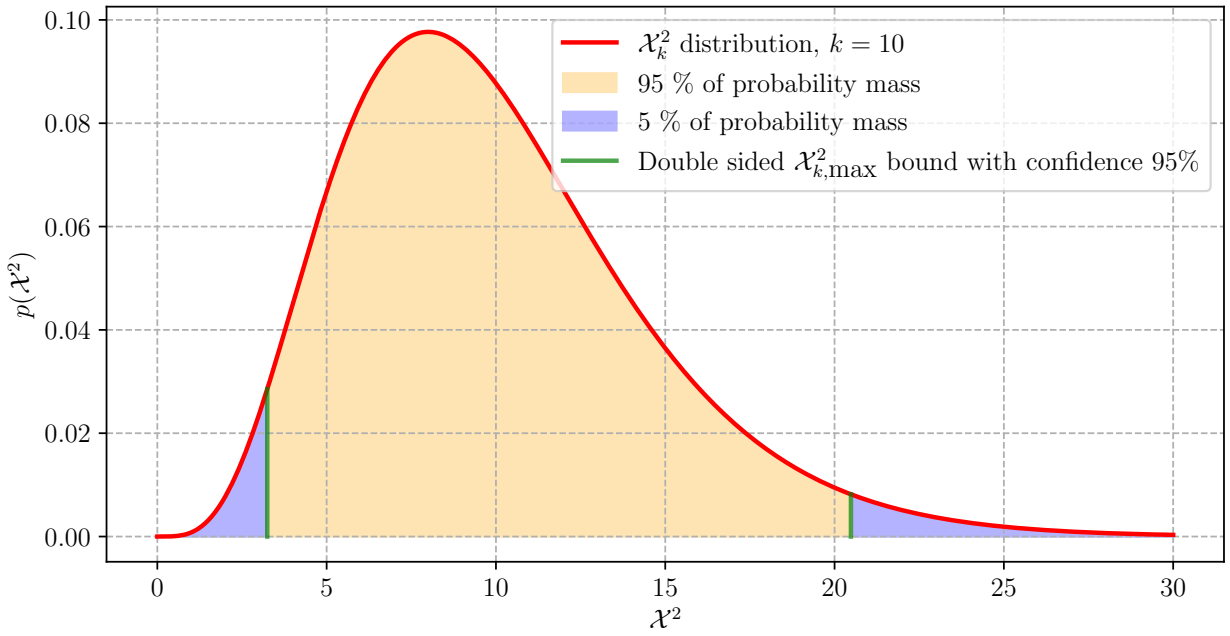

Figure 2: Illustration of double-sided χ^2_k squared test.

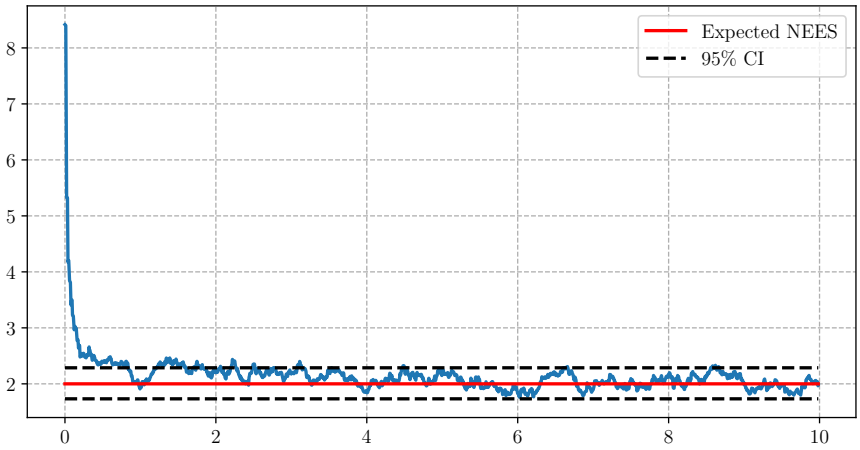

Figure 3: The average NEES for $N = 200$ Monte Carlo trials of an estimation problem with a single integrator process model and range measurements to anchors.

Mahalanobis distances for each trial, at a specific timestep i, is given by

$$d_i^2 = \sum_{n=0}^{N} \mathbf{e}_{i,n}^{\mathsf{T}} \mathbf{P}_{i,n}^{-1} \mathbf{e}_{i,n} \tag{26}$$

$$= \sum_{n=1}^{N} \mathbf{u}^{\mathsf{T}}\mathbf{u} \quad \mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{1}_{k\times k}) \tag{27}$$

$$= \sum_{j=1}^{kN} u_j^2 \quad u_j \sim \mathcal{N}(0,1), \tag{28}$$

and is thus a chi-squared distribution χ^2_{kN} with kN degrees of freedom. Therefore, for the upper and lower bounds, kN degrees of input is used as the input to the probability point function when computing the bounds. Furthermore, the squared Mahalanobis distance is typically normalized by the number of trials, in which case the corresponding bounds are then divided by the number of trials N.

The results for an example involving a two-dimensional position state with velocity and range measurements are illustrated in Fig. 3. The NEES is computed and averaged for $N = 200$ Monte Carlo runs. The number of degrees of freedom for computing the $\mathcal{X}_k^2$ distribution is given by $kN = 200 * 2 = 400$ and the resulting NEES is divided by $N = 200$ to yield the average NEES over all the Monte Carlo runs.

References

[1] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, "Estimation with Applications to Tracking and Navigation," New York, USA: John Wiley & Sons, Inc., 2001. (visited on 08/16/2023).